

Die Vergabe von DOI-Namen für Sozial- und Wirtschaftsdaten: Serviceleistungen der Registrierungsagentur da|ra

Hausstein, Brigitte

Veröffentlichungsversion / Published Version
Arbeitspapier / working paper

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Hausstein, B. (2012). *Die Vergabe von DOI-Namen für Sozial- und Wirtschaftsdaten: Serviceleistungen der Registrierungsagentur da|ra*. (RatSWD Working Paper Series, 193). Berlin: Rat für Sozial- und Wirtschaftsdaten (RatSWD). <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-427886>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

193

Die Vergabe von DOI-Namen für Sozial- und Wirtschaftsdaten

Serviceleistungen der Registrierungsagentur da|ra

Brigitte Hausstein

Januar 2012

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Working Paper Series des Rates für Sozial- und Wirtschaftsdaten (RatSWD)

Die *RatSWD Working Papers* Reihe startete Ende 2007. Seit 2009 werden in dieser Publikationsreihe nur noch konzeptionelle und historische Arbeiten, die sich mit der Gestaltung der statistischen Infrastruktur und der Forschungsinfrastruktur in den Sozial-, Verhaltens- und Wirtschaftswissenschaften beschäftigen, publiziert. Dies sind insbesondere Papiere zur Gestaltung der Amtlichen Statistik, der Ressortforschung und der akademisch getragenen Forschungsinfrastruktur sowie Beiträge, die Arbeit des RatSWD selbst betreffend. Auch Papiere, die sich auf die oben genannten Bereiche außerhalb Deutschlands und auf supranationale Aspekte beziehen, sind besonders willkommen.

RatSWD Working Papers sind nicht-exklusiv, d. h. einer Veröffentlichung an anderen Orten steht nichts im Wege. Alle Arbeiten können und sollen auch in fachlich, institutionell und örtlich spezialisierten Reihen erscheinen. Die *RatSWD Working Papers* können nicht über den Buchhandel, sondern nur online über den RatSWD bezogen werden.

Um nicht deutsch sprechenden Nutzer/innen die Arbeit mit der neuen Reihe zu erleichtern, sind auf den englischen Internetseiten der *RatSWD Working Papers* nur die englischsprachigen Papers zu finden, auf den deutschen Seiten werden alle Nummern der Reihe chronologisch geordnet aufgelistet.

Einige ursprünglich in der *RatSWD Working Papers* Reihe erschienenen empirischen Forschungsarbeiten sind ab 2009 in der RatSWD Research Notes Reihe zu finden.

Die Inhalte der *RatSWD Working Papers* stellen ausdrücklich die Meinung der jeweiligen Autor/innen dar und nicht die des RatSWD.

Herausgeber der RatSWD Working Paper Series:

Vorsitzender des RatSWD (2007/2008 Heike Solga; seit 2009 Gert G. Wagner)

Geschäftsführer des RatSWD (Denis Huschka)

Die Vergabe von DOI-Namen für Sozial- und Wirtschaftsdaten

Serviceleistungen der Registrierungsagentur da|ra

Brigitte Hausstein

GESIS – Leibniz-Institut für Sozialwissenschaften, Berlin

brigitte.hausstein(at)gesis.org

Abstract:

Das GESIS Leibniz-Institut für Sozialwissenschaften und das ZBW Leibniz-Informationszentrum für Wirtschaftswissenschaften betreiben in Kooperation mit DataCite, der internationalen Initiative zur Verbesserung des Zugangs zu Forschungsdaten, einen DOI-Registrierungsservice für Sozial- und Wirtschaftsdaten. Mit dieser Infrastruktur wird eine wichtige Voraussetzung für eine dauerhafte Identifizierung, Sicherung, Lokalisierung und schließlich eine verlässliche Zitierbarkeit von Forschungsdaten aus den Sozial- und Wirtschaftswissenschaften geschaffen.

Um die technischen und organisatorischen Lösungen für die Vergabe von DOI-Namen zu testen, führte GESIS 2010 ein Pilotprojekt für die Registrierung von sozialwissenschaftlichen Daten durch. Inzwischen sind 5200 Studien registriert und mehr als 7000 Metadatensätze in das Informationssystem aufgenommen worden.

Dieser Beitrag beschreibt die technische und organisatorische Implementierung der Registrierungsagentur da|ra und zeigt, wie in der Etablierungsphase des Projektes das bereits existierende DOI-Registrierungssystem ab 2012 auch für wirtschaftswissenschaftliche Forschungsdaten genutzt werden kann.

Keywords: Forschungsdaten, Datenzitation, Persistent Identifier, DOI-Namen

1. Hintergrund

Nicht nur der Zugang zu Literatur, sondern auch allgemein zugängliche und langfristig verfügbare Forschungsdaten sind für eine Wissenschaft, die exzellente Forschungsergebnisse erbringen will, essentiell. Die schnelle Entwicklung der digitalen Technologien und Netzwerke der letzten Jahre hat die Produktion, Verbreitung und Nutzung von Forschungsdaten jedoch radikal verändert. Neue Verfahren und Messinstrumente bringen wachsende, aber auch komplexere Datenmengen und -typen hervor. Entsprechend entwickelte Software-Tools helfen die Fülle der gesammelten Primärdaten zu verwalten, zu interpretieren und in Informations- und Wissenssammlungen zu transformieren. Das wichtigste und allseits präsente Forschungswerkzeug, das Internet, hat die Art und Weise, wie Daten und Informationen ausgetauscht und verfügbar gemacht werden, stark verändert (vgl. Uhlig 2003).

Während für Forschungspublikationen neben den traditionellen Angeboten der freie Zugang (*Open Access*) immer mehr zur gängigen Praxis wird, sind die Bemühungen hinsichtlich allgemein zugänglicher Datenpublikationen erst am Anfang. Obwohl grundsätzlich die Bereitschaft zur Weitergabe der Primärdaten existiert, scheitert dies oft an den fehlenden Kapazitäten, die für die Aufbereitung und Metadatenbeschreibung notwendig sind. Dies gilt auch für die Sozialwissenschaften, die im Vergleich zu anderen Disziplinen bereits eine ausgeprägte Kultur des „Data Sharings“ kennen.

Die Verbreitung der Forschungsergebnisse erfolgt fast ausschließlich noch über Publikationen in Fachzeitschriften. Die „Allianz der deutschen Wissenschaftsorganisationen“ hat jedoch Ende Juni 2010 in den „Grundsätze(n) zum Umgang mit Forschungsdaten“¹ eine Regelung für Primärdaten gefordert, um bei Wissenschaftlerinnen und Wissenschaftlern das Bewusstsein für den Handlungsbedarf und für den Nutzen von Primärdaten-Infrastrukturen zu schärfen. Von Seiten der Forschungsfinanzierer wird zunehmend gefordert, nicht nur die Forschungspublikationen sondern auch die entstandenen Primärdaten im Sinne von *Good Scientific Practice* öffentlich zugänglich zu machen. Daraus ergibt sich die besondere Bedeutung einer reinen Datenpublikation, mit allen Möglichkeiten der eindeutigen Identifikation und kompakten Zitierung, die für Textpublikationen bereits Standard sind.

Darüber hinaus zeigt sich in allen Wissenschaftsbereichen auch die Dringlichkeit der Kopplung von Datenarchivierung und wissenschaftlicher Literaturpublikation. Die Trennung von Forschungspublikation und zugrundeliegenden Daten erschwert die Evaluation der Publikation und schränkt die Nachvollziehbarkeit der dargestellten Ergebnisse ein. Bislang erfolgt die

1 <http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/grundsaeetze/>

Verbindung von Daten- und Forschungspublikation nur punktuell. Voraussetzung für die Verbindung von Forschungsprimärdaten und wissenschaftlicher Publikation ist jedoch neben der Langzeitarchivierung der Daten und entsprechender Qualitätssicherung, die Möglichkeit zur Publikation von Daten mit eindeutiger Identifizierbarkeit und Referenzierbarkeit.

2. Persistent Identifier und Forschungsdatenregistrierung

Ein Weg zur Lösung der geschilderten Problematik ist der Einsatz von speziellen Persistent Identifier. Deren Funktion entspricht in etwa einer ISBN-Nummer bei gedruckten Werken, die lediglich ein einziges Mal vergeben wird. Hinzu kommt die Unterscheidung zwischen dem Identifier und der Lokation eines Objekts, die es ermöglicht, das Objekt unabhängig von seinem Speicherort zu identifizieren. Dies unterscheidet die Persistent Identifier von einer *Universal Resource Locator (URL)*. Zur Sicherstellung der eindeutigen Vergabe und der Zuweisung von Kennung und Speicherort bedarf es eines automatisierten Dienstes. Jedem Persistent Identifier werden dabei Adressinformationen, z.B. ein URL zugewiesen. Von zentraler Bedeutung sind hier geeignete organisatorische Maßnahmen, die Verweise auf die tatsächlichen Speicherorte der Ressourcen aktuell halten. Programme können dann über einen sogenannten Resolverdienst den zitierten Persistent Identifier auflösen, so dass ein Zugang zu den zitierten Forschungsdaten möglich wird.

Es existieren mittlerweile für die Identifikation von elektronischen Textpublikationen diverse Systeme von Persistent Identifier, die technisch die Basis für einen Service auch zur Identifizierung von Daten leisten können: *Archival Research Key (ARK)*, *Digital Object Identifier (DOI®)*, *Handle*, *Library of Congress Control Number (LCCN)*, *Life Science Identifiers (LSID)*, *Persistent URL (PURL)*, *Uniform Resource Name (URN)* und weitere. Auf einen gemeinsamen Standard haben sich die verschiedenen Nutzergemeinden jedoch noch nicht geeinigt, da die Systeme im Prinzip gut ineinander überführbar sind. Um die langfristige Eignung zu beurteilen ist hier weniger die technische als die organisatorische Ausgestaltung relevant.

3. Das DOI[®]-System

Das DOI[®]-System² wurde von der *Association of American Publishers* entwickelt und wird gegenwärtig von der *International DOI Foundation (IDF)* verwaltet. Die IDF besteht seit 1998 und unterstützt die Rechteverwaltung für geistiges Eigentum in digitalen Netzwerken indem sie die Entwicklung und Verbreitung des DOI[®]-Systems als gemeinsame Infrastruktur für das *Content Management* fördert. Die IDF ist als *not-for-profit* Organisation registriert und wird von einem *Executive Board*, das von den Mitgliedern des IDFs gewählt wird, kontrolliert. Die Mitgliedschaft ist offen für alle Organisationen, die sich mit elektronischem Publizieren und den damit verbundenen Technologien beschäftigen.

Das DOI[®]-System ist ein verwaltetes System für die persistente Identifikation von Inhalten, die in digitalen Netzwerken angeboten werden. Es kann für die Identifizierung von physikalischen, digitalen oder anderen Objekten benutzt werden. Die Identifikatoren (DOI-Namen) führen direkt zum Speicherort des bezeichneten Objektes. Technisch basiert das DOI- System auf der vom CNRI³ entwickelten *Handle Technology*⁴. Es wird ergänzt durch ein Metadatenmodell, um die zum Objekt gehörenden Metadaten mit dem DOI-Namen zu verbinden. Auf der Basis der gemeinsamen Policy und technischen Infrastruktur der IDF wird das DOI[®]-System durch einen Zusammenschluss von Registrierungsagenturen umgesetzt. Dieses Lizenzmodell wird fälschlicher Weise mit einer kommerziellen Ausrichtung des DOI[®]-Systems verwechselt. Jede Registrierungsagentur kann jedoch über ihr eigenes Businessmodell für die Vergabe der DOI-Namen entscheiden.

DOI-Namen und Langzeitarchivierung

Das DOI[®]-System hat gute Aussichten auf Verbreitung und Langlebigkeit. Dies wird nicht zuletzt durch verbindliche Verträge zwischen Registrierungs-agentur und Nutzern erreicht, die eine längere Zeitperspektive versprechen. Insbesondere die internationalen wissenschaftlichen Fachverlage setzen fast durchgängig auf die Verwendung von DOI-Namen (vgl. Brase et al. 2009). Das DOI-System überzeugt daher auch für den Einsatz im Wissenschaftsbereich.

Die Verwendung von DOI-Namen im Rahmen der Langzeitarchivierung von digitalen Ressourcen bietet sich aus mehreren Gründen an. Neben der breiten Verwendung garantiert die überwachende Einrichtung der IDF die Einhaltung der Standards und die notwendige Persistenz. Der Identifier selbst kann aber keine dauerhafte Verfügbarkeit sicherstellen; es ist „nur“ eine technische Lösung, die Bestandteil jedes Langzeitarchivierungskonzeptes sein

² <http://www.doi.org>

³ Corporation for National Research Initiatives® (CNRI): <http://www.cnri.reston.va.us>

⁴ <http://www.handle.net>

sollte. Langzeitarchivierung impliziert die Sicherung über große Zeiträume. Keines der existierenden Persistent Identifier Systeme besitzt jedoch eine 100-prozentige Gewähr für Dauerhaftigkeit. Allerdings ist die Technik des Handle Systems so ausgelegt, dass eine Registrierungsagentur jederzeit komplett selbstständig die Auflösbarkeit ihrer DOI-Namen sicherstellen kann.

Struktur und Resolving eines DOI-Namens

Ein DOI-Name besteht genau wie ein *Handle* immer aus einem Präfix und einem Suffix, wobei beide durch einen Schrägstrich getrennt werden und das Präfix stets mit „10.“ beginnt (vgl. Abb.1). Das Präfix, das beispielsweise einem bestimmten Datenzentrum zugeordnet ist (in der Abb. 1: „3478“), ermöglicht die Bildung einer unbegrenzten Anzahl von DOI-Namen, indem auf der Basis des Präfixes und verschiedener Suffixe eine beliebige Reihe von eindeutigen Identifiern gebildet werden können.

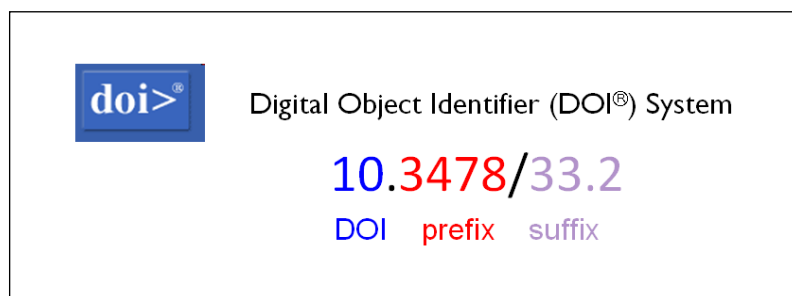


Abb. 1: Struktur eines DOI-Namens

Um einen DOI-Namen zur zugehörigen URL aufzulösen gibt es verschiedene Möglichkeiten, die alle auf dem zentral betriebenen Handle Server⁵ basieren. Zum einen kann er über das vom CNRI angebotene Resolverplug-in⁶ eingegeben und aktiviert werden. Eine andere Möglichkeit ist die Verwendung des Proxy Servers des DOI-Systems (<http://dx.doi.org/>) bzw. des Handle Systems (<http://hdl.handle.net/>). Die Eingabe des DOI-Namens zusammen mit der vorangestellten Zeichenkette <http://dx.doi.org/> in den Eingabeschlitz jedes beliebigen Browsers führt den Nutzer direkt zum Speicherort des Objektes bzw. zu einer Webseite, die dieses und die Zugangsbedingungen ausführlich beschreibt.

5 Dieser wird ab 2012 nicht mehr vom CNRI sondern von der AT&T, einer für die UNO in Genf tätigen Gesellschaft betreut.

6 <http://www.handle.net/resolver/>

4. DataCite

Seit 2010 sind GESIS und ZBW Mitglieder in DataCite. Damit wurde die Voraussetzung (DOI-Vergaberecht) für die Etablierung eines Registrierungsservices für Forschungsdaten auf der Basis von DOI-Namen geschaffen. DataCite⁷ ist ein 2009 in London gegründetes internationales Konsortium mit inzwischen sechzehn Mitgliedern aus zehn Ländern, die gemeinsam das Ziel verfolgen, die Akzeptanz von Forschungsdaten als eigenständige, zitierfähige wissenschaftliche Objekte zu fördern.

Während DataCite eine bei der IDF akkreditierte DOI-Registrierungsagentur ist, fungieren GESIS und ZBW als Vollmitglieder in DataCite als DOI-Allocation Agency.



Find, access, and reuse data

DataCite bietet neben der Mitgliedschaft in der IDF eine international abgestimmte Vorgehensweise bei technischen Lösungen, Standards, Best Practises und Workflows, sowie ein gemeinsames Metadatenschema und einen MetadataStore.

5. Registrierungsagentur da|ra

GESIS hat im Frühjahr 2010 ein Pilotprojekt unter dem Titel da|ra⁸ (Registrierungsagentur für sozialwissenschaftliche Daten) zur Etablierung eines Registrierungssystems für Forschungsdaten der Sozialforschung im deutschsprachigen Raum gestartet. Ziel war es, eine Infrastruktur zu entwickeln, die es ermöglicht, Forschungsdatenbestände mit DOI-Namen zu versehen und sie mit ihren Titeln, Themen, Autoren, Provenienzen, Methoden und Zugangsmöglichkeiten so umfassend wie möglich nachzuweisen sowie findbar und zitierbar zu machen.

da|ra

Registrierungsagentur für
Sozial- und Wirtschaftsdaten

In Kooperation mit DataCite wurde dazu mit der technischen Implementierung eines Registrierungstools begonnen und ein spezifisches über das

⁷ <http://www.datacite.org/>

⁸ <http://www.gesis.org/dara>

von DataCite hinausgehende Metadatenmodell entwickelt. Begonnen wurde mit der Registrierung der Forschungsdaten des GESIS-Datenarchivs um die technischen und organisatorischen Lösungen zu testen.

In einer zweiten Projektphase (bis Ende 2012) wird der Service weiteren Forschungsdatenzentren, Datenservicezentren und institutionalisierten Forschungsprojekten angeboten. Schwerpunkt in dieser Phase ist aber auch die Entwicklung des DOI-Registrierungsservices für Wirtschaftsdaten auf der Basis der vorhandenen technischen Lösung. Dazu wurde die Zusammenarbeit mit der ZBW aufgenommen. Dies lag nahe, da die Nutzergemeinschaften beider Infrastruktureinrichtungen aus eng verwandten Disziplinen stammen.

Die Einbeziehung weiterer Datenrepositorien und Datenkuratoren als Publikationsagenten ist für die Ausbauphase (ab 2013) geplant. In dieser Phase wird der Dienst in Abhängigkeit von der Entwicklung eines *Self-Archiving Services* bei GESIS auch Einzeldatenproduzenten aus dem sozialwissenschaftlichen Bereich angeboten. Über ein von der DFG gefördertes Projekt werden zusätzlich die Entwicklung von Standards und Guidelines für die DOI-Registrierung sowie die Entwicklung eines erweiterten Nachweissystems realisiert.

5.1 Organisation und Verwaltung der Registrierung

Die DOI-Registrierung erfolgt bei da|ra in enger Kooperation mit den datenhaltenden Organisationen, den so genannten Publikationsagenten, die für die Pflege und Speicherung der Forschungsdaten sowie für die Metadatenpflege zuständig sind. Die Datensätze verbleiben bei den Datenzentren, da|ra speichert die Metadaten und macht die registrierten Inhalte über eine Datenbank recherchierbar.

Die Rahmenbedingungen und Voraussetzungen für die DOI Registrierung sind in einer Policy festgehalten. Darüber hinaus werden der Workflow und sämtliche Verantwortlichkeiten im Registrierungsprozess in einem Service Level Agreement vereinbart. Hierunter fallen Fragen zur Qualitäts- und Persistenzsicherung (Daten und Metadaten), zu Urheberrechten, Versionierungen, zur Verfügbarkeit des Services sowie deren Funktionalitäten. Bei der Festlegung der Details werden die Best Practice Empfehlungen von DataCite berücksichtigt.

Während der Etablierungsphase von da|ra wird der Registrierungsservice kostenneutral angeboten. Nach Abschluss der Einführungsphase werden die Betreiber von da|ra prüfen, ob die Erhebung einer Gebühr erforderlich ist. Ziel ist es, den Service so zu gestalten, dass die Betreiberkosten auch langfristig auf einem niedrigen Niveau gehalten werden können und die Nutzung des Basisservices weiterhin kostenfrei bleibt.

5.2 Das Metadatenmodell

Neben dem technischen DOI-Registrierungsservice leistet da|ra auch die Übernahme von Metadatenbeschreibungen in sein Datenbanksystem. Damit wird eine differenzierte Suche aller registrierten Datensätze ermöglicht, eine inhaltliche Beschreibung zur Verfügung gestellt und die Voraussetzung zur einheitlichen Zitation der Daten geschaffen. Unter Berücksichtigung der besonderen Herausforderungen der Dokumentation digitaler Datenobjekte wurde für das Informationssystem ein eigenes Beschreibungsschema (vgl. Hausstein et al. 2011) entwickelt. Dieses basiert auf dem Metadatenschema des GESIS-Datenbestandskatalogs und wurde in Anlehnung an das DataCite-Metadatenschema⁹ erweitert. Durch das Mapping mit dem DDI Standard¹⁰ und durch die Verwendung kontrollierter Vokabulare ist die Interoperabilität der erfassten Ressourcen mit anderen Datenbeständen auch im internationalen Rahmen gewährleistet.

Ein wesentliches Element der Datenbeschreibung ist die Möglichkeit der Versionierung, die gerade im Bereich von Primärdaten für die eindeutige Referenzierung eines Datensatzes unerlässlich ist. Da jede Version mit einem eigenen DOI-Namen und eigenen Metadaten versehen ist, wird die eindeutige Bezugnahme auf den einer Analyse zugrundeliegenden Datensatz und damit z.B. auch die Verifizierung von Analyseergebnissen ermöglicht.

Ein Schwerpunkt der Vereinbarung mit dem Publikationsagenten stellt die Sicherstellung der Qualität der Metadaten dar. Bei der Registrierung neu anfallende Metadaten werden durch da|ra zunächst geprüft und bei Bedarf entsprechend dem da|ra-Metadatenschema nachbearbeitet. In der Folge ist der Publikationsagent verantwortlich für die Aktualität der Metadaten, insbesondere die korrekte Auflösung des DOI-Namens, während da|ra für die permanente technische Auflösbarkeit zuständig ist. Hierdurch wird die Persistenz des Angebotes gewährleistet.

5.3 Technische Implementation

Die technische Implementation des Registrierungstools erfolgt über eine serviceorientierte Architektur. Zentrale Komponenten, wie beispielsweise das Registrieren eines DOI-Namens oder das Indexieren, wurden in separate Services ausgelagert. Entsprechend verfügt auch das zentrale Informationssystem über Schnittstellen, die als Webservice angesprochen werden können.

Die Metadaten werden in einer Datenbank gemäß DDI Standard abgespeichert. Dadurch werden ein Im- und Export der Daten im XML-Format auf einfache Weise ermöglicht und ein Austausch mit den Publikationsagenten direkt unterstützt. Für die Funktionalität der Suche wird ein Indexierungsserver

⁹ Dieses basiert auf ISO 690-2, dem Dublin Core Metadata Element Set, Version 1.1 und dem DOI-Metadaten Kernel der IDF und liegt z.Z. in der Version 2.2 vor: <http://schema.datacite.org/>.

¹⁰ <http://www.ddialliance.org>

eingesetzt (SOLR). Die Oberfläche zum Editieren der Metadaten wird von Projektmitarbeitern oder den Publikationsagenten benutzt um die Metadaten zu pflegen. Die Edition wird durch die Verwendung von kontrollierten Vokabularen unterstützt, damit möglichst standardisierte Metadaten erstellt werden.

6. da|ra Services

Die Registrierungsagentur da|ra bietet eine vollständige Infrastruktur für die DOI-Registrierung und die Metadatenverwaltung. Die Betreiber von da|ra sehen sich als Dienstleister für Daten- und Forschungszentren, die ihre Primärdaten mit DOI-Namen registrieren wollen. Dabei kann es sich um Surveydaten, Aggregatdaten, Microdaten aber auch um qualitative Daten handeln.



Abb. 2: da|ra Services

Im Rahmen des Registrierungsprozesses erhält jeder Datensatz nach Übermittlung der erforderlichen Metadaten einen eindeutigen DOI-Namen, wobei Granularität und Namensgestaltung vom Publikationsagenten festgelegt werden können. Der DOI-Name kann dann sofort über einen Resolverdienst zur entsprechenden URL aufgelöst werden. Das Serviceangebot von da|ra (vgl. Abbildung 2) umfasst gegenwärtig die DOI-Namensregistrierung, eine DOI-Resolvermöglichkeit, ein umfangreiches Metadaten-beschreibungsschema, das Metadatenhandling in einer Datenbank sowie verschiedene Möglichkeiten zur Übertragung der Metadaten. Darüber hinaus können interessierte Nutzer über eine einfache und erweiterte Suche in den Metadatenbeständen recherchieren.

Das Angebot zu *impact factors and peer review* sowie die Verlinkungen von Daten und Publikationen befindet sich noch im Aufbau und wird nach Abschluss des DFG Projektes als Prototyp angeboten.

Gegenwärtig nutzen folgende sozialwissenschaftliche Forschungsdatenzentren und Datenarchive den da|ra Service:

- 1) Forschungsdatenzentrum des Sozio-oekonomischen Panels (FDZ-SOEP)
- 2) Forschungsdatenzentrum PsychData des Leibniz-Zentrums für Psychologische Information und Dokumentation (ZPID)
- 3) Forschungsdatenzentrum Deutscher Alterssurvey (FDZ-DEAS)
- 4) Projekt Nationales Bildungspanel (NEPS)
- 5) GESIS Datenarchiv für Sozialwissenschaften
- 6) Forschungsdatenzentrum "Internationale Umfrageprogramme" bei GESIS
- 7) Forschungsdatenzentrum ALLBUS bei GESIS
- 8) Forschungsdatenzentrum "Wahlen" bei GESIS.

Diese haben insgesamt ca. 5200 Studien/Datensätze registriert, wobei der größte Anteil beim GESIS Datenarchiv liegt. Zusätzlich wurden ca. 2400 Metadatensätze von der *iLibrary*¹¹ der OECD, die ihre Bestände über die Partnerschaft mit der DOI-Registrierungsagentur *crossref*¹² registriert, in das Informationssystem übernommen. Somit stehen mehr als 7500 Metadatensätze für Recherchen im da|ra Informationssystem zur Verfügung.

7. Ausblick

Im Rahmen der Etablierungsphase des Projektes arbeiten GESIS und ZBW an der Anpassung des Registrierungssystems um die Veränderungen zu berücksichtigen, die durch die technischen und organisatorischen Neuerungen bei DataCite sowie durch die Erweiterung des Anwendungsbereichs von da|ra entstanden sind. Das betrifft die Überarbeitung des Metadatenschemas, die entsprechende Anpassung der Datenbank, den Anschluss des Registrierungstools an den neugestalteten *MetaDataStore* von DataCite sowie das Update der da|ra Policy und des Service Level Agreements.

Im Einzelnen bedeutet dies eine Erweiterung des Metadatenschemas um die Spezifik der Forschungsdaten aus den Wirtschaftswissenschaften abbilden zu können. Gleichzeitig werden die Änderungen, die im Metadatenschema von DataCite vorgenommen wurden, berücksichtigt. Im Dezember 2011 wurde die Version 2.2.1. des da|ra Metadatenschemas verabschiedet. Dieses bildet die Grundlage für die Anpassung des technischen Systems, die noch bis Anfang 2012 andauern wird. In diesem Zusammenhang werden auch die *Add-on Services* entwickelt. Begonnen wird mit dem Angebot einer

¹¹ <http://www.oecd-ilibrary.org/>

¹² <http://www.crossref.org/>

facettierten Suche, einem *link checker* und einem *help desk* für die Publikationsagenten.

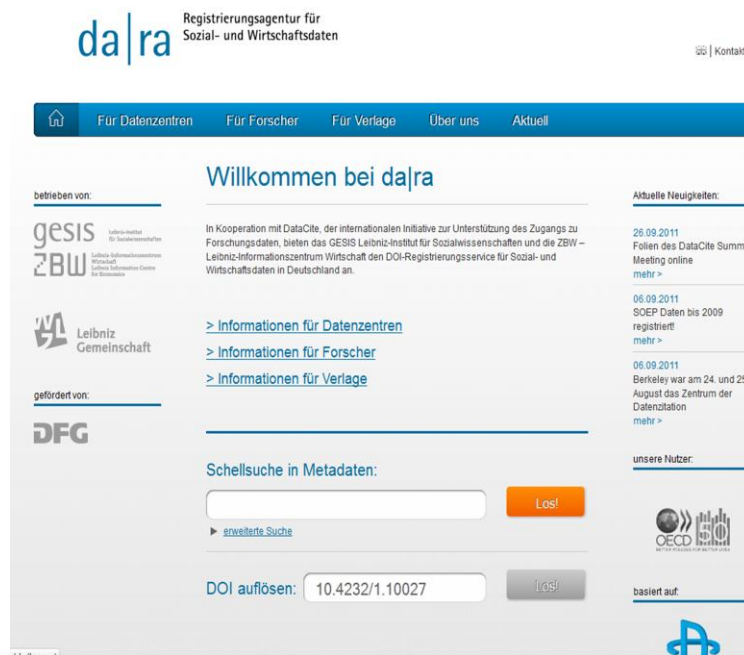


Abb. 3: Neuer Webauftritt von da|ra

Zukünftig werden GESIS und ZBW die Registrierungsagentur zwar gemeinsam betreiben, die Betreuung der Publikationsagenten wird aber entsprechend der disziplinären Zuordnung jeweils von GESIS oder ZBW erfolgen. Diese und andere daraus resultierenden Veränderungen im organisatorischen Ablauf mussten in der Policy und im Service Level Agreement ihren Niederschlag finden. Die da|ra Policy liegt nun in der Version 2.0 vor und berücksichtigt auch die von DataCite festgelegte cc0-Lizenzierung¹³ der Metadaten.

Das überarbeitete Angebot von da|ra wird ab Anfang 2012 über einen frisch gestalteten Webauftritt unter der neuen Domain **www.da-ra.de** angeboten. Besuchen Sie uns auf unserer Website und

Make Your Data Citable!

¹³ <http://creativecommons.org/publicdomain/zero/1.0/deed.de>

Literaturverzeichnis

- Allianz der deutschen Wissenschaftsorganisationen (2010): Grundsätze zum Umgang mit Forschungsdaten. 24. Juni 2010.
<http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/grundsaeetze/>
- Askitas, N. (2010): What Makes Persistent Identifiers Persistent? Working Paper No. 147. RatSWD Working Paper Series June 2010.
http://www.ratswd.de/download/RatSWD_WP_2010/RatSWD_WP_147.pdf
- Arbeitsgruppe Fachinformation (2009): Rahmenkonzept für die Fachinformation. Vorlage zur Sitzung des Ausschusses der Gemeinsamen Wissenschaftskonferenz des Bundes und der Länder (GWK) am 29.09.2009.
- Bellini, E., Cirinnà, C. et al. (2008): Persistent Identifiers distributed system for Cultural Heritage digital objects http://www.bl.uk/ipres2008/presentations_day2/38_Lunghi.pdf
- Bleuel, J. (2000): Zitation von Internet-Quellen (Citing of Internet sources). In: Hug, T. (Hrsg.): Wie kommt Wissenschaft zu Wissen? Band1: Einführung in das wissenschaftliche Arbeiten. Hohengehren: Schneider Verlag.
- Blue Ribbon Task Force (2010): Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf
- Brase, J. (2010): DataCite - A global registration agency for research data Working Paper No 149. RatSWD Working Paper Series July 2010.
- Brase J. et al. 2009: Approach for a global joint registration agency for research data. Information Services & Use. In: Neuroth H. et al. (Hrsg.): nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung
- Brase, J. & Klump, J. (2007): Zitierfähige Datensätze. Primärdaten-Management durch DOIs. In: Wissenschaftskommunikation der Zukunft. 4. Konferenz der Zentralbibliothek. http://dc110dmz.gfz-potsdam.de/contenido/std-doi/upload/pdf/Brase_Wisskomm_2007.pdf
- DFG (1998): Vorschläge zur Sicherung guter wissenschaftlicher Praxis: Empfehlungen der Kommission „Selbstkontrolle in der Wissenschaft“, Denkschrift der Deutschen Forschungsgemeinschaft.
http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_0198.pdf
- Diepenbroek, M. & Grobe, H. (2007): PANGAEA® als vernetztes Verlags- und Bibliothekssystem für wissenschaftliche Daten. In: Wissenschaftskommunikation der Zukunft. 4. Konferenz der Zentralbibliothek. http://www.fz-juelich.de/zb/datapool/page/1000/Diepenbroek_Abstract.pdf
- Dittert, N., Diepenbroek, M. & Grobe, H. (2002): Data and information management for the CMTT synthesis. Manuskript <http://epic.awi.de/Publications/Dit2002b.pdf>
- GESIS Report 4/10. Der Aktuelle Informationsdienst für die Sozialwissenschaften. GESIS - Leibniz Institute für Sozialwissenschaften Mannheim September 2010.
http://www.gesis.org/fileadmin/upload/institut/presse/gesis_report/gesis_report_10_04.pdf
- Hausstein, B. & Zenk-Möltgen, W. (2011): daJa – Ein Service der GESIS für die Zitation sozialwissenschaftlicher Daten. In: Schomberg, S., Leggewie, C. & Puschmann, C. (Hrsg.) Digitale Wissenschaft. Stand und Entwicklung digital vernetzter Forschung in Deutschland. 20/21. September 2011 Köln. Beiträge der Tagung. www.hbz-nrw.de/Tagung_Digitale_Wissenschaft.pdf
- Hausstein, B., Zenk-Möltgen, W., Wilde, A. & Schleinstein, N.: daJa Metadatenschema. Version 1.0 GESIS-Working Papers 2011|14. GESIS – Leibniz-Institut für Sozialwissenschaften 2011. DOI:10.4232/10.mdsdoc.1.0
- Klump, J., Bertelmann, R. et al. (2006): Data publication in the open access initiative. Data Science Journal, Volume 5, Number 79.
- Neuroth, H., Oßwald, A., Scheffel, R. Strathmann, S. Huth, K. (2007): nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.3. nestor – Network of Expertise in Long-Term Storage of Digital Resources. Göttingen. <http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:0008-2010071949>
- Metadata for the Publication and Citation of Scientific Primary Data (Version 3.0)
http://www.icdp-online.org/contenido/std-doi/upload/pdf/STD_metadata_kernel_v3.pdf

- Paskin, N. (2000): Digital Object Identifier: implementing a standard digital identifier as the key to effective digital rights management. The International DOI Foundation Kidlington, Oxfordshire, United Kingdom http://www.doi.org/doi_presentations/aprilpaper.pdf
- Parsons, M. A., Duerr, R. & Minster, J.-B.: Data Citation and Peer Review. In: Eos, Transactions, American Geophysical Union, Vol. 91, No. 34, 24 August 2010, PAGES 297–304
<http://aurora.gmu.edu/spaceweather/images/2010EO340001.pdf>
- Riding the Wave: How Europe can gain from the rising tide of scientific data. Final report of the High Level Expert Group on Scientific Data. A submission to the European Commission. October 2010. <http://goo.gl/WrxO>
- Reilly, S. (2010): Digital Object Repository Server: A Component of the Digital Object Architecture. D-Lib Magazine January/February 2010. Volume 16, Number ½
<http://www.dlib.org/dlib/january10/reilly/01reilly.print.html>
- Uhlir, P. F. (2003): Discussion Framework. In Esanu, J. M. & Uhlir, P. F. (Hrsg.): The Role of Scientific and Technical Data and Information in the Public Domain. Washington DC: The National Academies Press.
- Uhlir, P. F. & Schroder, P. (2008): Chapter 8 - Open Data for Global Science. In Fitzgerald, B.: Legal Framework for e-Research: Realising the Potential. Sydney: Sydney University Press Law Books 39 (2008) 188.
<http://www.austlii.edu.au/au/journals/SydUPLawBk/2008/39.html>

Internet

- Allianz der deutschen Wissenschaftsorganisationen
<http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/grundsatzet/>
- California Digital Library: ARK <http://www.cdlib.org/inside/diglib/ark/>
- Corporation for National Research Initiatives® (CNRI) <http://www.cnri.reston.va.us>
- crossref <http://www.crossref.org>
- da|ra <http://www.gesis.org/dara> <http://www.da-ra.de>
- DataCite <http://www.datacite.org>
- Data Documentation Initiative <http://www.ddialliance.org/>
- Digital Object Identifier (DOI®) System <http://www.doi.org/>
- Forschungsdatenzentrum des Sozio-oekonomischen Panels (FDZ-SOEP)
- Forschungsdatenzentrum Deutscher Alterssurvey (FDZ-DEAS)
- Forschungsdatenzentrum PsychData des Leibniz-Zentrums für Psychologische Information und Dokumentation (ZPID)
- Forschungsdatenzentrum "Internationale Umfrageprogramme" bei GESIS
- Forschungsdatenzentrum ALLBUS bei GESIS
- Forschungsdatenzentrum "Wahlen" bei GESIS
- GESIS - Leibniz Institut für Sozialwissenschaften (Mannheim, Köln, Berlin)
<http://www.gesis.org>
- Homepage der Technischen Informationsbibliothek (TIB) Hannover zu eigener DOI-Registrierungsagentur <http://www.tib-hannover.de/de/die-tib/doi-registrierungsagentur/>
- Homepage der Technischen Informationsbibliothek (TIB) Hannover zu Projekt CODATA
<http://www.tib-hannover.de/de/die-tib/projekte/codata/>
- Library of Congress: Structure of the LC Control Number
http://www.loc.gov/marc/lccn_structure.html
- Life Sciences Identifiers (LSID) <http://lsids.sourceforge.net/>
- OECD iLibrary <http://www.oecd-ilibrary.org>
- Persistent Uniform Resource Locator (PURL) <http://purl.oclc.org/docs/index.html>
- Projekt Nationales Bildungspanel (NEPS)
- Publikation und Zitierfähigkeit wissenschaftlicher Primärdaten http://www.std-doi.de/front_content.php
- The Handle System® <http://www.handle.net>
- Uniform Resource Name (URN) Syntax <http://tools.ietf.org/html/rfc2141>
- ZBW <http://www.zbw.eu>